DOCUMENT RESUME

ED 448 383 CG 030 577

AUTHOR Hoover, Heather V. A.; Braver, Sanford L.; Wolchik, Sharlene A.; Sandler, Irwin N.

TITLE Teachers' Ratings of Children's Classroom Behavior: Time of Year Effects?

PUB DATE 2000-08-00

NOTE 14p.; Poster session presented at the Annual Conference of the American Psychological Association (108th, Washington, D.C., August 4-8, 2000).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Behavior Problems; *Behavior Rating Scales; Classroom Observation Techniques; Divorce; Elementary Education; Elementary School Students; *Personality Assessment; *School Psychologists; *Student Behavior

ABSTRACT

School psychologists rely heavily upon teachers' evaluation of children's classroom behavior as part of their assessment and intervention processes. However, some evidence suggests that teachers' ratings may fluctuate during the academic year, with higher levels being reported later in the year. This study assessed the comparability of different teachers' ratings of children's externalizing behavior from fall and spring intervention cohorts. Teachers of children from divorced families (N=240, ages 9 to 12) completed pretest, posttest, and six-month follow-up ratings. Two separate ANOVAs revealed that changes in mean levels were not significantly different between the fall and spring cohorts for either the pretest/posttest or pretest/follow-up comparison. Ratings obtained from children's current teachers within the first month of school were compared to those completed by the children's teachers from the previous school year. Pretest levels of externalizing behavior were higher when reported retrospectively by the previous teachers. However, neither the previous teachers' nor the current teachers' fall ratings were significantly different from the spring teachers' pretest ratings. Thus, school psychologists may elect to obtain ratings from either the previous or current teacher early in the fall of a new academic year in the assessment of children's behavior. (Contains 2 tables and 12 references.) (JDM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 448 383

TEACHERS' RATINGS OF CHILDREN'S CLASSROOM BEHAVIOR: TIME OF YEAR EFFECTS?

Poster presented at the 108th Annual Meeting of the American Psychological Association (Washington, DC) in August, 2000 by:

Heather V. A. Hoover, M.A.
Sanford L. Braver, Ph.D.
Sharlene A. Wolchik, Ph.D.
& Irwin N. Sandler, Ph.D.

Arizona State University
Department of Psychology

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

For more information, please contact:
Heather V. A. Hoover, M.A.
Arizona State University
Clinical Psychology Ph.D. Program
Department of Psychology
PO Box 871104
Tempe, AZ
85287-1104

Phone (480) 965-3326
Fax (480) 965-8544
Email: heather.hoover@asu.edu

H. HOOVER

CG030577

BEST COPY AVAILABLE

Abstract

Teachers' reports may be affected by measurement and interpretation difficulties associated with the potential differences between ratings completed during the fall and spring, particularly if different teachers provide the ratings. This study assessed the comparability of teachers' ratings of children's externalizing behavior from fall and spring intervention cohorts. Teachers of 240 children of divorce aged 9 to 12 years completed pretest, posttest, and six-month follow-up ratings. Children were randomly assigned to one of three treatment conditions (groups for their mothers only, separate groups for mothers and children, or a bibliotherapy control condition). Two separate 3 X 2 X 2 ANOVAs revealed the changes in mean levels were not significantly different between the fall and spring cohorts for either the pretest/posttest or pretest/follow-up comparison. When different teachers completed the ratings, the pretest/posttest and pretest/follow-up correlations were remarkably high over 9 months (*rr* = .55 to .79). When the pretests were obtained from the current teachers, the pretest/follow-up correlation was higher in the fall than in the spring. A second aim compared ratings obtained from children's current teachers within the first month of the school year vs. those completed by the children's teachers from the previous school year. Pretest levels of externalizing behavior were higher when reported retrospectively by the previous teachers. However, neither the previous teachers' nor the current teachers' fall ratings were significantly different from the spring teachers' pretest ratings. Thus, school psychologists may elect to obtain ratings from either the previous or current teacher early in the fall of a new academic year. Limitations of this study and recommendations for future research are discussed.

Introduction

School psychologists rely heavily upon teachers' reports of children's classroom behavior. Although this information is quintessential to the assessment and intervention processes, the validity of this data may depend upon the time of the academic year that the data are collected. Some evidence suggests that teachers' ratings of students' behavior may fluctuate during the academic year. For example, within the first few weeks of the academic year, teachers do not tend to report high levels of behavioral problems (Taylor-Greene et al., 1997). Therefore, ratings obtained early in the academic year may show little variability. On the other hand, greater variability and higher levels of behavior problems may plausibly be reported later in the academic year.

When a student is referred for an evaluation or is re-evaluated early in the academic year, school psychologists must decide whether to obtain behavior ratings from the student's current or previous teacher, or both. It is generally recommended that a teacher have a student in his or her class for at least four weeks before providing behavior ratings (e.g., Conners, 1973; Merrell, 1994; Quay & Peterson, 1983; Reynolds & Kamphaus, 1992; Walker, 1983). The authors of the standardized rating instrument used in the present study, the Teacher-Child Rating Scale (TCRS), specify four to six weeks as ample time for teachers to effectively complete ratings on social, behavioral, and academic behaviors (Hightower et al., 1986; Primary Mental Health Project, 1995).

As a result of their current teachers' unfamiliarity with students early in the academic year, school psychologists may obtain ratings from the teacher whose classroom the child was in the previous academic year. However, this strategy is problematic in that the reports are retrospective, do not account for the child's maturation over the summer, and reflect the child's behavior in a previous setting.

When teachers' ratings of a student's behavior are needed within the first few weeks of a academic year (e.g., as part of an evaluation or as a baseline measurement prior to intervention), school psychologists must select one teacher or reconcile possible discrepancies between multiple teachers' ratings of the same student. Currently, empirical evidence to guide this decision-making is lacking.

A related issue involves the use of teachers' reports to evaluate a student's progress in behavioral domains. When behavior ratings are used to monitor a student's response to an intervention, teachers may be asked to report on the student's behavior in the fall and again in the spring semester of an academic year. If the intervention period spans more than one academic year, different teachers are likely to provide ratings of the same student's behavior at different time points. The validity of the ratings may be compromised by inter-rater unreliability. On the other hand, when the same teacher provides pre-intervention and post-intervention ratings, the ratings may be less sensitive to detecting change because the teacher' s perceptions of the child's behavior are resistant to change (Bryk & Raudenbush, 1992; Raudenbush, 1984). Although assessments are intended to obtain reasonably objective measurements of children's behavior, the teachers and classroom settings may affect the behavior ratings at different assessment points. Thus, teachers' reports may be particularly affected by measurement and interpretation difficulties associated with the potential differences between fall and spring behavior ratings.

Among the issues that could lead to systematic differences between fall and spring ratings are (1) the length of time a child has been in a particular class with a particular teacher, (2) whether the ratings are completed by the same or different teachers at different assessment points, and (3) whether the sample of observed behavior changes in a different setting. At present, little is known about semester-specific effects as they affect the evaluation of interventions for school-age children. This study assessed the comparability of teachers' ratings of children's externalizing behavior from fall and spring intervention cohorts.

Participants

Teachers of 240 children aged nine to twelve years completed pretest, posttest, and six-month follow-up ratings for four different cohorts (two in the fall, two in the spring) over a two-year period. There were 107 subjects in the fall cohorts and 133 in the spring. Forty-eight percent of the children were female. Eighty-eight percent of the children were Caucasian. At the time of the pretest, the average age of the children on whom teachers completed behavior ratings was 10.35 years. The grade level of the children ranged from three to seven, with a mode of fourth grade and a mean of 4.85.

The children on whom the teachers completed ratings were part of a preventive intervention study of divorced families ("New Beginnings", Wolchik, Sandler, West, & Anderson, 1997). The families were recruited to participate in the study via a multiple step process that was initiated by obtaining court records of divorce decrees that were granted within two years of the start of each intervention group.

Eligibility criteria

(1) The divorce decree was granted within the past two years;
(2) The mother was the custodial/primary residential parent;
(3) Neither the custodial parent nor any child residing with the custodial parent was currently in treatment for psychological problems;
(4) The custodial parent had not remarried, did not have a live-in boyfriend, and did not plan to remarry during the time period of the intervention trial;
(5) The custodial arrangement was expected to remain stable during the trial;
(6) The family resided within a one-hour drive of the location of the intervention group program;
(7) The custodial parent and the child being targeted by the intervention were fluent in English;
(8) The target child was not placed in a special education program for the mentally handicapped or learning disabled;
and (9) One or more children between nine and twelve years of age resided primarily with their mother.

Method

Measure

Teacher-Child Rating Scale (T-CRS; Hightower et al., 1986; Primary Mental Health Project, 1995). The Teacher-Child Rating Scale includes a six-item Acting Out subscale that taps externalizing behavior problems. The T-CRS has demonstrated adequate reliability as well as concurrent and discriminant validity to differentiate referred from non-referred children (Hightower et al., 1986).

In the present study, the internal consistency of the Acting Out subscale was excellent. The reliability coefficients (Cronbach's alphas) for the Acting Out subscale exceeded .90 on the pretest, posttest, and follow-up ratings.

Data Collection

In the current investigation, the pretest assessments were obtained within the four weeks prior to the start of the intervention. Posttest ratings were gathered three months after the pretest and within the four weeks immediately following the completion of the intervention. Follow-up assessments occurred six months after the post-intervention assessments. The ratings for spring cohorts were gathered as follows: pretest in late February; posttest in May; and follow-up in November. For fall cohorts, the ratings were collected as follows: pretest in early September; posttest in December; and follow-up in early June. For fall cohorts, pretest data were collected from both the previous and current teachers of each child in the study.

Fall Pretest Completed by the Current Teachers

The fall teachers who completed the pretest ratings on the fall cohorts had known the children for an average of three weeks. The range was wide, with some teachers reporting having had a child in class for only one week, while others reported knowing a child for as many as six weeks. These ratings were completed in early September.

Treatment Conditions

After completing the pretest, participants attended an orientation session during which they were randomly assigned to one of three conditions: (1) mother-only group intervention (N=81), (2) concurrent but separate mother and child group interventions (N=83), and (3) a control condition (N=76). The control condition consisted of self-study, take-home reading materials that mothers were at the time when mothers were randomly assigned into this condition. The intervention subjects in conditions (1) and (2) participated in semester-long groups aimed at intervening with the putative mediators of divorce-related events for children (Wolchik et al., 1997).

Design

Dividing the sample into the two experimental (mother plus child and mother only) conditions and one control (self study/bibliotherapy) group resulted in a 3 X 2 X 2 analysis of variance (ANOVA) with one repeated measures (within-subjects) factor and two between-subjects factors. This design was replicated as another 3 X 2 X 2 ANOVA for the pretest/follow-up scenario.

The repeated measures factor was the time each rating was collected on each subject (i.e., pretest, posttest, and follow-up). The cohort in which a child was a subject was a non-random, between-subjects factor. The treatment condition, either experimental or control was a random, between-subjects factor.

For each of the comparisons between fall and spring cohorts, correlation coefficients, means, and differences between means were computed. Mean externalizing scores were examined to assess the substantive question regarding differences between cohorts

as a function of time of year. Correlation coefficients between scores at two assessment points (pretest/posttest and pretest/follow-up) were calculated on the within-subjects factor to address the statistical question of sensitivity to detecting treatment effects.

All of the comparisons of correlation coefficients described herein account for the three treatment conditions, such that each computation actually refers to partial correlation coefficients. Partial correlation coefficients were computed between pretest and posttest scores for each of the three situations: spring (same teacher), fall (same teacher), and fall (different teachers).

Results

Two separate 3 X 2 X 2 ANOVAs revealed the changes in mean levels were not significantly different between the fall and spring cohorts for either the pretest/posttest or pretest/follow-up comparison. When different teachers completed the ratings, the pretest/posttest and pretest/follow-up correlations demonstrated stability over nine months (r = .55 to .79). (In a meta-analysis of thirteen studies that used two teachers as informants on children's behavior, Achenbach, McConaughy, and Howell (1987) found a mean r = .64 on ratings completed by different teachers for the same child.) When the pretests were obtained from the current teachers in the fall, the pretest/follow-up correlation was higher in the fall cohorts than in the spring cohorts.

Pretest Mean Levels: Fall Current vs. Previous Teachers

Ratings obtained from children's current teachers within the first month of the academic year were compared to those completed by the children's teachers from the previous academic year. Both sets of these fall pretest ratings were obtained at about the same time, early in the fall.

Main effects were found when looking at pretest means between the ratings obtained from the current teachers and the previous teachers in the fall cohorts. The previous teachers reported very highly significantly greater symptomatology than did the current teachers, F (1, 84) = 12.80, $p < .001$. As was the case when fall and spring pretest ratings were compared, there was not a significant interaction between treatment condition (i.e., group) and the effect of which teacher completed the pretest rating. This finding fits the expectation of randomization into treatment conditions.

Conclusions

Regardless of the underlying reasons, it is clear that teachers' ratings of children's externalizing behavior are quite consistent, even when different teachers provide behavior ratings over a period of nine months. The correlations between ratings over time were remarkably high, even when different teachers completed the questionnaires. Ratings completed nine months apart by different raters correlated from .55 to .74. Achenbach et al. (1987) reported a mean correlation of .64 between different teachers' ratings, typically obtained at about the same time. There were no significant differences between the fall and spring cohorts' partial correlation coefficients of the pretest/posttest. As such, school psychologists may elect to obtain behavior ratings from either or both the previous and current teacher early in the fall of the new academic year.

Limitations of this study include the lack of information about whether the stability of teachers' ratings is actually problematic with regard to detecting treatment effects. Also, the generalization of the findings should be approached with caution for a number of reasons. First, the instrument used in this study, the Teacher-Child Rating Scale (T-CRS, Hightower et al., 1986) lacks normative data for children beyond third grade. Second, the present sample was comprised entirely of children of divorce. Although the inclusion of a control condition should safeguard against problems with making appropriate comparisons, it is unknown whether the same results would be obtained with a different sample.

Future research should examine teachers' ratings of internalizing behaviors and verify the validity of teachers' ratings compared to behavioral observations. Further research could elucidate the mechanisms through which teachers' ratings become stable and determine if teachers and classroom environments themselves become contexts that promote stability in the way children are perceived.

Table 1

Design of the Study

Fall cohorts

Pre	Post	FU
T1B & T2B	T2A	T2F
Sept.	Dec.	June

Spring cohorts

Pre	Post	FU
T3B	T3A	T4F
Feb.	May	Nov.

Note. T = Teacher, and is followed by a number assigned to differentiate whether the teacher is the same or different teacher as the one who completed a rating at another time point. The letters after the numbers indicate when the rating took place:

B = Before, A = After, and F = Follow-up.

Table 2

Partial Correlations

	Pre/Post			Pre/Follow-up		
		pr	n		pr	n
Fall (previous pre)	(T1B/T2A)	.63a*	81	(T1B/T2F)	.60a	76
Fall (current pre)	(T2B/T2A)	.79b*	81	(T2B/T2F)	.74b	76
Spring	(T3B/T3A)	.75ab	113	(T3B/T4F)	.55a	113

Note. *Fall current* indicates the pretest data came from each child's current teacher at the time of the rating. *Fall previous* indicates that the pretest rating was obtained from each child's teacher from the previous school year.

Note. Coefficients in the same column sharing the same superscript are not significantly different from one another.
* $p < .01$; all other values with different superscript letters are significantly different at $p < .05$.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situation specificity. Psychological Bulletin, 101 (2), 213-232.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods (pp. 161-168).

Conners, C. K. (1973). Rating scales for use in drug studies with children. In Psychopharmacology Bulletin: Pharmacatherapy with children. Washington, DC: Government Printing Office.

Hightower, A. D., Work, W. A., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., & Rohrbeck, C. A. (1986). The Teacher-Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. School Psychology Review, 15, 393-409.

Merrell, K. W. (1994). Assessment of behavioral, social, and emotional problems: Direct and objective methods for use with children and adolescents. New York: Longman.

Primary Mental Health Project. (1995). Screening and evaluation measures and forms: Teacher-Child Rating Scale. Rochester, NY: Author.

Quay, H. C., & Peterson, D. R. (1983). Interim manual for the Revised Behavior Problem Checklist. Privately printed.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from eighteen experiments. Journal of Educational Psychology, 76 (1), 85-97.

Reynolds, C. R., & Kamphaus, R. W. (1992). Behavioral Assessment System for Children: Teacher Rating Scales. Circle Pines, MN: American Guidance Service.

Taylor-Greene, S., Brown, D., Nelson, L., Longton, J., Gassman, T., Cohen, J., Swartz, J., Horner, R. H., Sugai, G., & Hall, S. (1997). School-wide behavioral support: Starting the year off right. Journal of Behavioral Education, 7 (1), 99-112.

Walker, H. M. (1983). Walker Problem Behavior Identification Checklist - manual. Los Angeles: Western Psychological Services.

Wolchik, S. A., Sandler, I. N., West, S. G., & Anderson, E. (1997). Children of divorce: Six-year followup of preventive efforts. Application submitted for grant renewal.